Structural genomics and signaling domains

James H. Hurley, D. Eric Anderson, Bridgette Beach, Bertram Canagarajah, Yew Seng Jonathan Ho, Eudora Jones, Greg Miller, Saurav Misra, Matt Pearson, Layla Saidi, Silke Suer, Ray Trievel and Yosuke Tsujishita

Many novel signal transduction domains are being identified in the wake of genome sequencing projects and improved sensitivity in homology-detection techniques. The functions of these domains are being discovered by hypothesis-driven experiments and structural genomics approaches. This article reviews the recent highlights of research on modular signaling domains, and the relative contributions and limitations of the various approaches being used.

The concept of modular signaling domains has been at the center of signal transduction research for the past 15 or so years. The discovery of domain families such as SH2, SH3, PDZ, PH, C1 and C2, and the understanding of their properties, have contributed immeasurably to our knowledge of signaling pathways. The importance of the domain concept is reflected in the fact that each one of the domains mentioned above occurs in hundreds of different signaling proteins. Once the function of a particular domain from one protein is well understood, powerful and testable inferences can be made as to the function of the many other proteins that contain that domain. Thus, domain information provides the simplest and most powerful conceptual bridge between otherwise overwhelmingly vast and complex sequence data. Because of this, great effort has gone into understanding the structures and functions of these domains, leading to our ability to rationalize, and in some cases predict, subtle differences in specificity.

The number of known signaling domain families has expanded rapidly in the past few years. This trend is driven by the value of the biological information to be gleaned, and is made possible by data from genome sequencing and by high-sensitivity sequence-homology detection. In 1997, all the known intracellular signaling domains were described in the pages of *TiBS* [1], a total of 37 domains. The signaling domain database SMART (Ref. [2] and http://smart.embl-heidelberg.de/) now contains 150 entries categorized as 'intracellular signaling'. Newly discovered domain families mined from homology searches of worldwide sequence databases are similar to raw ore in that the most valuable content must still be 'smelted' out of these sequences by other techniques. Bioinformatics-based discoveries of new signaling domain families sometimes define biochemical function in a clear-cut

way, as when one or more family members correspond to protein fragments whose activity has been previously characterized. At least as often, the identification of a new protein family poses the new question: what is (are) the function(s) of the FITB (fill in the blank) domain? The volume of domain discoveries is so great, and this question is raised so frequently, that it has spawned a new enterprise dedicated to its answer.

The problem of identifying the function of a protein starting from its sequence is central to structural and functional genomics, but it has been cast into particularly sharp focus when applied to signaling domains. For this review, we have chosen several examples of domain families whose structures and functions have recently been uncovered (Table 1). These cases illustrate broader trends in the synergy between the traditional hypothesis-driven paradigm of biochemical research and the more recent discovery-driven paradigm.

The examples in the table also illustrate how structural biology has had a dramatically increased presence in the early stages of understanding the function of newly discovered domains. The systematic structure determination of signaling domains as a class fits at least one of the definitions of structural genomics [3]. This brings us to the title of this article: 'Structural genomics and signaling domains'. The examples described in this article from our own group and from that of Shapiro were the result of explicitly taking a structural genomics approach (Fig. 1). The authors of the other studies do not describe their work explicitly in structural genomics terms, but we argue that approaches being taken, the targets chosen and the collective outcome of these studies are not so different from what might have been produced by a coordinated effort. As large-scale structural genomics initiatives begin in earnest, recent experiences with signaling domains offer hints about what might be in store as these initiatives move from early demonstration projects into areas with wide-ranging impacts on fundamental questions in biology and molecular medicine.

Tubby

The tubby-like protein (TULP) family first came to light because, in mice, the mutation of the gene for

James H. Hurley*
D. Eric Anderson
Bridgette Beach
Bertram Canagarajah
Yew Seng Jonathan Ho
Eudora Jones
Greg Miller
Saurav Misra
Matt Pearson
Layla Saidi
Silke Suer
Ray Trievel
Yosuke Tsujishita
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0580, USA.
*e-mail: jh8e@nih.gov

Table 1. Signaling domains with recently described structures[a]

Domain	Function[b]	Biochemical function discovered from[c]:	SMART sized[d]	Actual size[e]	Structure
DEP	Unknown	NA	75	94	3-helix bundle plus β-hairpin arm
TULP core	PtdIns(4,5)P_2-regulated transcription factor	Structural genomics	NA	263	Novel fold with a β-barrel filled by C-terminal α helix, basic groove for DNA binding
START	Lipid monomer binding and transport	Structural genomics	206	229	Unclosed β-barrel capped by a C-terminal helix, hollowed out core with a hydrophobic tunnel for lipid binding, similar fold to mammalian PITP and to plant allergens
ENTH	PtdIns(4,5)P_2 and protein–protein interactions	Hypothesis-based	137	149	Helical superhelix similar to VHS
VHS	Endocytic signal sequence binding	Hypothesis-based	136	153	Helical superhelix similar to ENTH
PX	Binds PtdIns(3)P, other phosphoinositides	Hypothesis-based	118	143	Novel α+β fold
PB1	Protein–protein interaction with 'PC' motif in small G proteins and others	Hypothesis-based	NA	79	Similar to Ras-binding domain of Raf
GAF	Binds cGMP, chromophore, other small molecules	Inferred from larger proteins	150	180	Similar to PAS, another sensory and signaling domain
IPP5C	Phosphoinositide 5-phosphatase	Inferred from larger proteins	299	336	Similar to DNAseI and DNA repair enzymes such as APE1

[a]Abbreviations: NA, not applicable; PtdIns(3)P, phosphatidylinositol (3)-monophosphate; PtdIns(4,5)P_2, phosphatidylinositol (4,5)-bisphosphate.
[b]'Function' refers to the best-known function(s) for the domain group, and is not inclusive of all cases.
[c]The approach indicated is that judged by the authors to be the most important single contributor to revealing the biochemical (as opposed to the cellular or genetic) function. In general, multiple approaches contributed. 'Inferred from larger proteins' is distinct from 'hypothesis-based' in that at least some of the functions of the former were established before the domain was identified as a conserved signaling motif.
[d]The predicted size (number of amino acid residues) of the domain from SMART alignments, choosing a particular domain for which a structure has been determined.
[e]The actual size (number of amino acid residues) of the domain as solved by X-ray crystallography or nuclear magnetic resonance, including essential extensions where present. These extensions do not appear to be part of the core fold of the domain in all cases, yet their inclusion is often essential to obtaining folded and functional protein. The differences between the predicted and actual sizes illustrate that substantial experimental effort is required to determine the correct boundaries of a predicted domain.

which the family is named, leads to maturity-onset obesity [4,5]. The TULP core domain comprises the conserved C-terminal portion of these proteins. Genetic and sequence information did not suggest a biochemical mechanism for the role of the TULP domain in obesity. Instead, structure determination proved to be the key to understanding the function of the TULP domain [6]. The TULP story is one of the most interesting contributions of structural genomics, combined with other discovery-driven assays, to enhancing our understanding of signaling mechanisms.
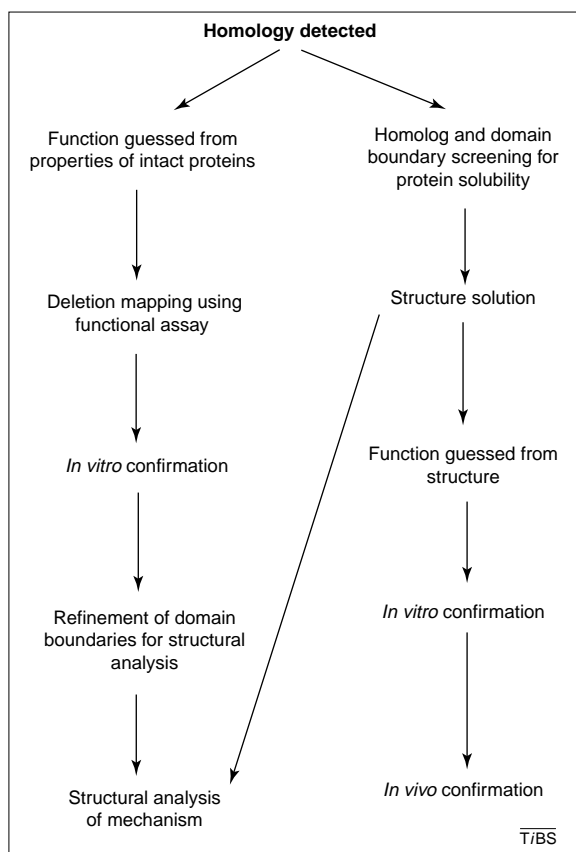
The tubby–TULP core domain has a novel fold comprising a 12-stranded β-barrel that completely encircles a hydrophobic α helix that runs the length of the barrel [6]. The clue to the function of TULP came from a large curved basic surface with an appropriate size, shape and charge for binding DNA. Structure determination led to the hypothesis that the TULP core domain was the DNA-binding domain of a transcription factor, which was borne out by functional assays [6]. However, the question of the upstream regulation of tubby remained unanswered. This time, an answer was suggested by using the signal trap assay and by monitoring subcellular localization using green fluorescent protein (GFP) fusions. The tubby–TULP core domain localizes initially to the plasma membrane, but slowly dissociates and enters the nucleus [7]. This observation led to the hypothesis that lipid turnover at the plasma membrane drove membrane desorption, and again the hypothesis was confirmed by further experimentation, including structure determination of the TULP–PtdIns(4,5)P_2 headgroup complex.

The DEP domain

The DEP (dishevelled–eglin–pleckstrin homology) domain is a widespread motif found in proteins involved in wnt signaling, regulators of G-protein signaling (RGS) proteins, pleckstrin and other signaling proteins [8]. Binding partners are not known for any DEP domain, but a role has been established for the Dishevelled DEP domain in wnt signaling based on mutations that interfere with biological function (reviewed in Ref. [9]). The DEP domains of several proteins have a membrane-targeting function [10,11], although the molecular mechanism of targeting is unknown.

If the tubby story represents the best-case scenario for structural genomics-based elucidation of function, the recent nuclear magnetic resonance (NMR) structural analysis of the Dishevelled DEP domain is perhaps a more typical illustration of what can and cannot be gleaned about a signaling domain from its three-dimensional structure alone [9]. The structure of this 100 amino acid domain consists of a three-helix bundle, a β-hairpin and two short C-terminal β strands. The analysis did not reveal any structural similarities to known structures that might provide clues as to the function of this domain. However, molecular surface features of Dishevelled-DEP did provide two clues as to the biochemical function of this domain. The first clue came from the observation of a cluster of seven basic residues that formed a flat patch on one side of the domain, which is typical of acidic phospholipid membrane-binding sites. This suggests that the membrane-targeting mechanism of DEP domains might involve direct binding of DEP domains to the negatively charged and roughly

Fig. 1. Two simplified paradigms for discovering the function of a newly identified domain, a function-based and hypothesis-driven approach (left), and a discovery-driven approach (right). The shunt from 'structure solution' to 'structural analysis of mechanism' illustrates just one aspect of the many levels of interplay between the different approaches, which are often being executed simultaneously by different laboratories.

Homology detected

Function guessed from properties of intact proteins

↓

Deletion mapping using functional assay

↓

In vitro confirmation

↓

Refinement of domain boundaries for structural analysis

↓

Structural analysis of mechanism

Homolog and domain boundary screening for protein solubility

↓

Structure solution

↓

Function guessed from structure

↓

In vitro confirmation

↓

In vivo confirmation

T*i*BS

planar surface of phospholipid bilayers in the cell. The second clue came from a mutation of a Lys residue that was already known to abrogate the function of Dishevelled in wnt signaling (reviewed in Ref. [9]). Lys434 is on a different face of the domain to the putative membrane-interacting face. This suggests that there are at least two functionally important interaction sites on the domain. The story of DEP domain function will remain incomplete, however, until binding partners of this domain are identified.

START: lipid transporters, signaling proteins and more
START domains are found in a surprisingly diverse collection of proteins, including known and putative lipid transporters, transcription factors, enzymes of lipid metabolism, and signaling proteins [12]. The START domain is named after the steroidogenic acute regulatory (StAR) protein. The StAR protein is crucial for steroid hormone production because it is essential for the delivery of cholesterol to the inner membrane (IM) of mitochondria, where the first enzymatic reaction of steroidogenesis takes place. Mutations within the StAR-START domain cause congenital lipoid adrenal hyperplasia (reviewed in Ref. [12]). Although the cellular role of StAR is well-known, the biochemical mechanism whereby cholesterol is delivered to the IM has been unclear. The recent crystal structure of the START domain of a cousin of StAR, MLN64, revealed a hollowed-out protein containing a hydrophobic tunnel big enough to bind one molecule of cholesterol and completely exclude it

from solvent [13] (Fig. 2). The structure-based hypothesis was confirmed by direct binding studies on the StAR and MLN64 START domains [13]. This suggests that the START domain functions in lipid transport by binding lipid monomers and sequestering them from solvent to deliver them across the aqueous compartments of the cell.

Taking it to the ENTH degree
The ENTH (Epsin N-terminal homology) [14] and VHS (Vps27, Hrs, STAM) [15] domains occur at the N-termini of proteins involved in intracellular trafficking. Crystal structures of Epsin-ENTH [16], Hrs-VHS [17] and Tom1-VHS [18] revealed that these two domains have very similar structures consisting of right-handed, eight-helical superhelices. The structures described in the first round of studies focused attention on, but fell short of answering, the question of the biochemical function of the ENTH and VHS domains. In the case of ENTH, the breakthrough came from structural and functional studies of the extended ENTH domains of two other endocytic proteins, AP180 [19] and CALM [20]. These proteins are regulated by PtdIns(4,5)P_2, and studies of the N-terminal domains showed this regulation to be mediated by their ENTH domains. Crystal structures of the phosphoinositide complexes followed, revealing the binding sites, and also showing that these N-terminal domains are essentially an expanded version of the ENTH domain. Concurrently, the Epsin-ENTH was shown to weakly bind PtdIns(4,5)P_2 at a site near but not identical to the AP180 and CALM sites [21]. This discovery might not have been convincing taken in isolation, but makes sense in context with the AP180 and CALM results. The AP180 and CALM ENTH domains bind PtdIns(4,5)P_2 with moderate affinity at a site with several basic residues. The developing picture is that the ENTH domain family, similar to the pleckstrin homology (PH) domain family, is a collection of phosphoinositide-binding domains with widely varying affinity and specificity.

High fidelity: VHS domains
Despite the structural similarities of VHS and ENTH domains, they appear to serve distinct functions. Again, biological hypotheses drove the biochemical discoveries. A family of VHS domain-containing proteins called the GGAs (Golgi-localized γ-adaptin ear-domain-containing) was recently reported to direct intracellular trafficking in the endosomal–lysosomal pathway. With the other domains accounting for functions in Arf and clathrin binding, the VHS was the only remaining orphan domain in the GGAs. Known orphan cargo proteins then became candidates for interactions with the GGA VHS domains. This approach revealed that the function of the GGA-VHS domain is to recognize an acidic-cluster-dileucine signal within the cytoplasmic C-terminal tails of membrane proteins that are trafficked from the *trans* Golgi network to
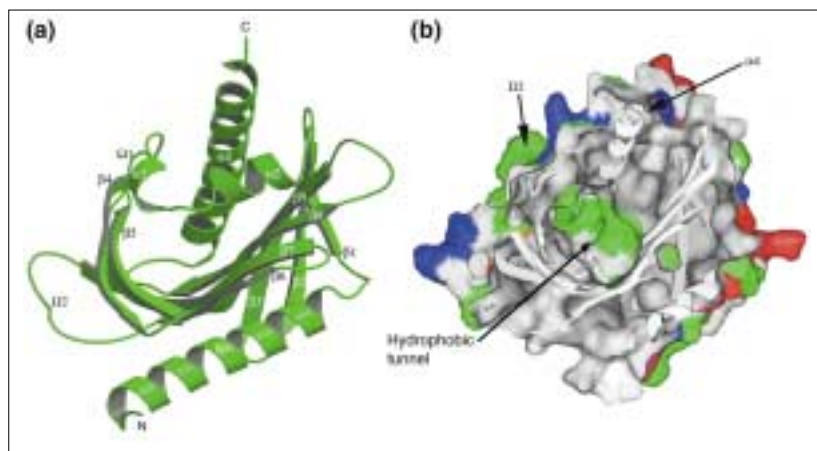
Fig. 2. (a) The START domain fold, an unclosed β-barrel capped a C-terminal helix. (b) Sagittal slice through the molecular surface of the MLN64 START domain shown in the same perspective as in (a). The surface cutaway reveals a mostly hydrophobic tunnel running through the center of the protein. The observation of a hydrophobic tunnel in the middle of this domain was the central observation leading to the proposal that START domain proteins bind and transport lipid monomers. The surface is colored blue (basic), red (acidic), green (hydrophobic) and white (uncharged polar). The interior of the protein is gray. Figure is adapted, with permission, from Ref. [13].

endosomes [22–25]. The structures of GGA-VHS domains bound to signal peptides were solved in quick succession, and revealed the signal peptide binding sites (S. Misra *et al.*, unpublished).

PX

PX domains attracted attention in two waves. The first wave hit when their presence was noted in a large number of signaling proteins, especially the NADPH oxidase phox subunits for which the domain is named [26]. This was followed, after several years of latency, by a small tsunami of attention when four groups simultaneously reported that certain PX domains are cellular receptors for phosphatidylinositol 3-phosphate [PtdIns(3)P] [27–31]. Various other PX domains appear to bind other phosphoinositides [30,32]. The intellectual basis for the discovery resembled the case for ENTH: several PX-domain-containing proteins – Vam7p, SNX3, and the p40 and p47 subunits of the NADPH oxidase complex – were known to translocate in response to PI 3-kinase signaling. The PX domain became a natural candidate for the PtdIns(3)P-responsive module within these proteins, and from there it was a matter of pinning down the details. The NMR structure of the PX domain of p47[phox] was determined before the phosphoinositide-binding function of the PX domain had been established [33] . This study focused attention on the interaction of the PX domain with the SH3 domain of the same protein. Such intraprotein, interdomain interactions are undoubtedly important for allosteric regulation, although that is not usually what is construed when 'the function' of a domain is discussed. Fortunately, the second structure of a PX domain, following on the heels of the discovery of the function, was an X-ray study of the complex of the p40[phox]-PX with a short-chain PtdIns(3)P [34].

PB1

The PB1 motif is a newly described domain involved in interacting with the small G protein Cdc42p [35]. The trend towards the compression of events in bioinformatics, structure and function, was highlighted by two back-to-back reports on the discovery, function and structure of the PB1 domain [35,36]. The function of the Bem1p-PB1 domain was established in a straightforward way, by deletion mapping of the determinants for the already-known interaction with Cdc42p. The NMR structure of this domain [36] revealed similarity to the Ras-association domain of Raf, consistent with its function in the binding of small G proteins.

GAF

GAF domains [37] are among the most widespread of all signaling domains, found in 826 proteins in the SMART database. Cyclic GMP (cGMP)-regulated cyclic nucleotide phosphodiesterases (PDEs) contain GAF domains, which are the allosteric binding sites for cGMP. Hundreds of other signaling proteins also contain GAF domains, including essential plant signaling enzymes such as the photosensing phytochromes and the ethylene receptor, and a vast array of microbial signaling and sensory proteins. Thus, the functions of a handful of GAF domains are known, including the cGMP-binding GAF domains of the PDEs, and the chromophore-binding GAF domains of the phytochromes. Remarkably, the functions of most GAF domains remain unknown, and no structure of a GAF domain was known until last year.

The crystal structure of a GAF domain from a yeast protein of unknown function, YKG9, revealed a close structural similarity to another very widespread class of signaling and sensory domain, the PAS (Per Arnt Sim) domain [38]. Three PAS domain structures have been solved, and two out of the three bind a chromophore or a heme in a distinctive buried pocket on one side of the central β sheet. The GAF domain structure revealed an unusual buried pocket that coincides structurally with the heme- or chromophore-binding pocket of the PAS domain. Thus, the GAF and PAS domains together form a structurally, and almost certainly evolutionarily, related family of small-molecule-binding domains. What the structure does not reveal is the nature of the small molecule that binds in the pocket, either of the YKG9 GAF domain, or of other GAF domains. On occasion, structural genomics has succeeded in identifying small molecules and cofactors that bind to proteins of previously unknown function, as these ligands can be present in the host cells used for recombinant protein expression. This sort of strategic serendipity is most likely to occur when the expressing host cell is similar to the organism in which the protein occurs naturally, and cannot be counted upon when eukaryotic protein domains are expressed in a prokaryotic host.

The most pressing question for understanding the biological function of the large majority of GAF domain proteins is to determine the identities of the small-molecule ligands of these GAF domains. This remains a challenging objective, as no general method for doing this exists.

Inositol polyphosphate 5-phosphatases

The inositol polyphosphate 5-phosphatases have been intensively studied over the past decade because of their profound importance in an enormous range of cellular processes [39]. These enzymes, and their catalytic (IPP5C) domains, might seem to be in odd company among the domains described above, most of which were only very recently discovered and studied. Despite efforts by many groups over the years, no crystal structure of any well-characterized IPP5C domain has been obtained. The use of a structural genomics-inspired tactic, database searching for previously unstudied homologs that might be susceptible to crystallization, led to the cloning, characterization, and crystal structure determination of a new member of this family, a previously unnamed and uncharacterized protein from *S. pombe* that is now known as SPsynaptojanin [40]. Although the function of this domain was well-established, the structure led to new insights into the catalytic mechanism and substrate specificity, an example of the interplay between structural genomic and traditional structural biology approaches.

Conclusions

The studies described in this article illustrate some of the insights that have been obtained into a range of biological processes by approaches that focus on protein domains. It is worth considering the relative contributions of structural and functional approaches, and of hypothesis-driven versus discovery-driven paradigms. Different approaches represent a continuum between these poles, and it is the interplay between the approaches that is most revealing.

The tubby study is a vivid illustration of the power of a discovery-driven and structural genomic approach to function. The discovery that tubby is a transcription factor was driven by the observation of a large basic region on the structure that literally 'looked like' a DNA-binding site. Subsequent hypothesis-driven functional assays were essential to confirm this idea. The second tubby breakthrough was propelled by a different type of unbiased discovery-driven assay using GFP fusions. Similar to tubby, involvement of the StAR protein was implicated by genetics both in disease and in regulation of normal physiology, yet its biochemical function was unclear. The clues to function were more extensive than for tubby, but again, it was the crystal structure that was pivotal in terms of suggesting mechanism.

The ENTH, VHS and PX domain stories illustrate the undiminished power of a good, old-fashioned hypothesis. In the ENTH and PX cases, full-length proteins containing these domains were known to be involved in PtdIns(4,5)P_2- and PtdIns(3)P-dependent processes, respectively. The discovery of the function of the VHS domain had to wait for the discovery of a new class of VHS-domain-containing protein, the GGAs.

The first structures of ENTH, VHS, PX and DEP domains were determined before the elucidation of their biological functions, and were not particularly enlightening in a functional sense. What they did provide, at least for ENTH, VHS and PX (the function of the DEP domain remained unknown) was a powerful impetus for a second round of much more informative structural studies carried out in the full light of the known function. The second wave followed so closely on the heels of the first reports of known function partly because of the groundwork laid by the initial wave. Although structure might not always reveal function by itself, in the absence of pre-existing biochemical work, the structure of a novel domain casts what was heretofore a purely bioinformatic construction, into physical and chemical 'reality' for the first time.

What makes some structures more revealing about function than others? Bacterial structural genomics has had some notable success in predicting function from structure. Many of these successes have involved enzymes, in which fold similarity and conserved catalytic geometry provide powerful insights. Fold similarity is a powerful predictor of common function in these cases, but where ligand-binding domains are concerned, it might be much less useful. For example, the structural similarity between the ENTH and VHS domains did not prove to be a major factor in our understanding of their function. In many cases, bacterial enzymes expressed in a bacterial host are purified in complex with relevant cofactors. When a eukaryotic protein, to take YKG9-GAF as an example, is expressed in a prokaryotic host, there is no guarantee that the relevant cofactor or modification will be present.

In eukaryotes, where regulatory and signaling proteins outnumber metabolic proteins, identification of protein, nucleic acid, membrane lipid and small-molecule ligands is probably important in more cases than is identifying a novel enzyme activity. The primary function of many eukaryotic signaling domains is their binding to other proteins. For these cases, other scaleable approaches such as the yeast two-hybrid and pulldown or proteomics methods will be more suitable. It is often possible to spot a nucleic acid, phospholipid membrane, or hydrophobic small-molecule ligand-binding site by inspection of a structure, and these situations favor structural genomics. In summary, the question posed at the

outset of this article: 'what is the function of the FITB domain?', is difficult enough to answer to warrant the implementation of all available approaches.

Domains will probably represent the bulk of protein targets in eukaryotic structural genomics because so many intact eukaryotic proteins are large enough to be challenging to express and crystallize. The crucial difference between studying novel domains and intact proteins is that the correct boundaries need to be established for each new domain. In the past, the function of a protein domain was usually established before the structural work began. Functional analysis usually involves deletion mapping that produces at least a rough starting point for expression of protein for structural studies. Without the benefit of a functional assay, physical–chemical properties of the domain (typically meaning its solubility) are the only basis for the assay.

This significantly increases the effort involved, but not to a prohibitive degree. It is encouraging that so many domain structures (e.g. the first structures for ENTH, VHS, PX and DEP) are being produced where the solubility of the recombinant protein is the sole guide to the choice of boundaries.

If genome sequences are the 'parts list of life', one could argue that the sum total of signaling domains, together with their binding partners and post-translational modifications, represent a 'parts list for signal transduction'. It is gratifying to see the convergence of bioinformatic, structural and functional techniques, and the interplay between hypothesis- and discovery-driven paradigms. As attention in signal transduction shifts towards network models, the accumulation of domain data should contribute to quantitative and predictive models for cell signaling.

References

1 Bork, P. *et al.* (1997) Cytoplasmic signalling domains: the next generation. *Trends Biochem. Sci.* 22, 296–298

2 Schultz, J. *et al.* (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28, 231–234

3 Terwilliger, T.C. *et al.* (1998) Class-directed structure determination: foundation for a protein structure initiative. *Protein Sci.* 7, 1851–1856

4 Kleyn, P.W. *et al.* (1996) Identification and characterization of the mouse obesity gene tubby: a member of a novel gene family. *Cell* 85, 281–290

5 NobenTrauth, K. *et al.* (1996) A candidate gene for the mouse mutation tubby. *Nature* 380, 534–538

6 Boggon, T.J. *et al.* (1999) Implication of tubby proteins as transcription factors by structure-based functional analysis. *Science* 286, 2119–2125

7 Santagata, S. *et al.* (2001) G-protein signaling through tubby proteins. *Science* 292, 2041–2050

8 Ponting, C.P. and Bork, P. (1996) Pleckstrin's repeat performance: a novel domain in G-protein signaling? *Trends Biochem. Sci.* 21, 245–246

9 Wong, H.C. *et al.* (2000) Structural basis of the recognition of the Dishevelled DEP domain in the Wnt signaling pathway. *Nat. Struct. Biol.* 7, 1178–1184

10 Axelrod, J.D. *et al.* (1998) Differential recruitment of Dishevelled provides signaling specificity in the planar cell polarity and Wingless signaling pathways. *Genes Dev.* 12, 2610–2622

11 Koelle, M.R. and Horvitz, H.R. (1996) EGL-10 regulates G protein signaling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell* 84, 115–125

12 Ponting, C.P. and Aravind, L. (1999) START: a lipid binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* 24, 130–132

13 Tsujishita, Y. and Hurley, J.H. (2000) Structure and lipid transport mechanism of a StAR-related domain. *Nat. Struct. Biol.* 7, 408–414

14 Kay, B.K. *et al.* (1999) Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery. *Protein Sci.* 8, 435–438

15 Lohi, O. and Lehto, V.P. (1998) VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. *FEBS Lett.* 440, 255–257

16 Hyman, J. *et al.* (2000) Epsin 1 undergoes nucleocytosolic shuttling and its Eps15 interactor NH2-terminal homology (ENTH) domain, structurally similar to armadillo and HEAT repeats, interacts with the transcription factor promyelocytic leukemia Zn2+ finger protein (PLZF). *J. Cell Biol.* 149, 537–546

17 Mao, Y.X. *et al.* (2000) Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* 100, 447–456

18 Misra, S. *et al.* (2000) Structure of the VHS domain of human Tom1 (target of myb 1): insights into interactions with proteins and membranes. *Biochemistry* 39, 11282–11290

19 Mao, Y.X. *et al.* (2001) A novel all helix fold of the AP180 amino-terminal domain for phosphoinositide binding and clathrin assembly in synaptic vesicle endocytosis. *Cell* 104, 433–440

20 Ford, M.G.J. *et al.* (2001) Simultaneous binding of PtdIns(4,5)P_2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* 291, 1051–1055

21 Itoh, T. *et al.* (2001) Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* 291, 1047–1051

22 Nielsen, M.S. *et al.* (2001) The sortilin cytoplasmic tail conveys Golgi–endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* 20, 2180–2190

23 Puertollano, R. *et al.* (2001) Sorting of mannose 6-phosphate receptors mediated by the GGAs. *Science* 292, 1712–1716

24 Zhu, Y.X. *et al.* (2001) Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor. *Science* 292, 1716–1718

25 Takatsu, H. *et al.* (2001) GGA proteins interact with acidic dileucine sequences within the cytoplasmic domains of sorting receptors through their VHS domains. *J. Biol. Chem.* 276, 28541–28545

26 Ponting, C.P. (1996) Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3-kinases: binding partners of SH3 domains? *Protein Sci.* 5, 2353–2357

27 Wishart, M.J. *et al.* (2001) Phoxy lipids: revealing PX domains as phosphoinositide binding modules. *Cell* 105, 817–820

28 Cheever, M.L. *et al.* (2001) Phox domain interaction with PtdIns(3)*P* targets the Vam7 t-SNARE to vacuole membranes. *Nat. Cell Biol.* 3, 613–618

29 Xu, Y. *et al.* (2001) SNX3 regulates endosomal function through its PX-domain-mediated interaction with PtdIns(3)*P*. *Nat. Cell Biol.* 3, 658–666

30 Kanai, F. *et al.* (2001) The PX domains of p47phox and p40phox bind to lipid products of PI(3)K. *Nat. Cell Biol.* 3, 675–678

31 Ellson, C.D. *et al.* (2001) PtdIns(3)*P* regulates the neutrophil oxidase complex by binding to the PX domain of p40phox. *Nat. Cell Biol.* 3, 679–682

32 Song, X. *et al.* (2001) Phox homology domains specifically bind phosphatidylinositol phosphates. *Biochemistry* 40, 8940–8944

33 Hiroaki, H. *et al.* (2001) Solution structure of the PX domain, a target of the SH3 domain. *Nat. Struct. Biol.* 8, 526–530

34 Bravo, J. *et al.* (2001) The crystal structure of the PX domain from p40phox bound to phosphatidylinositol 3-phosphate. *Mol. Cell* 8, 829–839

35 Ito, T. *et al.* (2001) Novel modular domain PB1 recognizes PC motif to mediate functional protein–protein interactions. *EMBO J.* 20, 3938–3946

36 Terasawa, H. *et al.* (2001) Structure and ligand recognition of the PB1 domain: a novel protein module binding to the PC motif. *EMBO J.* 20, 3947–3956

37 Aravind, L. and Ponting, C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.* 22, 458–459

38 Ho, Y.S.J. *et al.* (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.* 19, 5288–5299

39 Majerus, P.W. *et al.* (1999) The role of phosphatases in inositol signaling reactions. *J. Biol. Chem.* 274, 10669–10672

40 Tsujishita, Y. *et al.* (2001) Specificity determinants in phosphoinositide dephosphorylation: crystal structure of an archetypal inositol polyphosphate 5-phosphatase. *Cell* 105, 379–389